

Building cyclic master surgery schedules with leveled resulting bed occupancy: A case study

Jeroen Beliën

Erik Demeulemeester

Brecht Cardoen

Katholieke Universiteit Leuven

Faculty of Economics and Applied Economics

Department DSIM: Decision Sciences & Information Management

Research Center for Operations Management

Naamsestraat 69, B-3000 Leuven, Belgium

email: jeroen.belien@econ.kuleuven.be

email: erik.demeulemeester@econ.kuleuven.be

email: brecht.cardoen@econ.kuleuven.be

phone: 0032 (0)16326972

Abstract

This paper describes a case study in which a number of optimization algorithms are applied in order to build a cyclic master surgery schedule with leveled resulting bed occupancy. The study starts from detailed information on all elective surgery interventions during a 1-year period in a medium-sized Belgian hospital. For each surgeon-hospitalization unit combination multinomial distribution functions are derived for both the number of operated patients per operating room block and the length of stay of each operated patient. These distribution functions serve as the input for the algorithms. Leveling is achieved by either mixed integer programming techniques involving the solution of a min-max optimization problem and a quadratic optimization problem, or a simulated annealing heuristic that minimizes the total probability of bed shortage or, alternatively, the total expected bed shortage.

Keywords: Operating room scheduling, case study, mixed integer programming, quadratic programming, simulated annealing

1 Introduction and literature review

The purpose of this paper is to present a real-life application of the theoretic models proposed in Beliën and Demeulemeester (2005). These models enable us to build a cyclic master surgery schedule for which the resulting bed occupancy is leveled as much as possible and for which performance measures as the daily expected bed occupancy, the variance on this occupancy, the expected bed shortage and the probability of a shortage on each day can be predicted. Multinomial distribution functions are assumed for both the number of patients per operating room block and the length of stay (LOS) of each operated patient. The models applied in this case study are slightly extended implementations of the ones presented in Beliën and Demeulemeester (2005). The most important extension includes that more than one hospitalization unit is considered, leading to probability distributions for each surgeon-hospitalization unit combination. Also, block sizes may vary or, in other words, room allocations to surgeons can have variable durations. This extension has some consequences for the simulated annealing approach for which we have added a corresponding neighborhood move (see further). Finally, a

real-life constraint is added that prevents individual surgeons from being scheduled in different rooms at the same time.

The operations research literature on operating room scheduling can be structured using the three stages that can be distinguished in developing operating room schedules. In the first stage, often called case mix planning, it is decided how the available operating room time is divided over the different surgeons (or surgical groups). This stage takes place on the strategic level of hospital management as it determines for which ailments capacity will be preserved for a long time horizon. Case mix planning problems have been studied by amongst others Hughes and Soliman (1985), Rifai and Pecenka (1989), Robbins and Tuntiwongbiboon (1989) and Blake and Carter (2002) and (2003).

The second stage involves the development of a master surgery schedule. From the first stage it is already known *how much* operating room time each surgeon obtains. In the second stage, it is determined *when* the surgeons obtain operating room time in the total operating room theatre. This stage can be situated on the tactical level of hospital management as master surgery schedules are usually built for periods ranging from six months to three years. Few papers have dealt with this stage of operating room scheduling. Blake et al. (2002) propose an integer programming model for building a cyclic schedule that allocates to each surgical group a number of operating room hours as close as possible to its target operating room hours, while at the same time keeping the schedule as simple (repetitive) as possible (see also Blake and Donald, 2002).

The third stage involves the detailed planning of each elective case. This stage has more of an operational focus as it occurs on a daily basis and includes operational scheduling decisions like assigning the cases to operating rooms, determining the order and the start and end times of the cases (e.g., Weiss, 1990; Ozkarahan, 1995 and 2000), the reservation of specialized equipment etc. Sometimes, this third stage is integrated with longer term scheduling. For instance, Guinet and Chaabane (2003) and Jebali et al. (2006) propose a two-step approach for operating theatre planning. The first step involves assigning patient interventions to operating rooms on a medium term

horizon. The second step entails the daily rescheduling of the patient interventions in order to integrate more characteristics for human and material resource synchronization.

Several objectives have gained attention in the literature on operating room scheduling so far. Many authors have developed and tested algorithms for maximizing the operating room utilization (see, e.g., Dexter et al., 1999; Dexter and Traub, 2002). Other work has focussed on minimizing the operating room staffing costs (e.g., Dexter et al., 2000). Many studies have focussed on increasing the punctuality of the schedule realized (e.g., Lapierre et al., 1999; Dexter et al., 2001; Marcon et al., 2003).

The management of resources is often considered a crucial issue in operating room scheduling. The models applied in this paper aim at leveling the bed occupancy. The relation between bed occupancy and the surgery schedule has already been studied by many authors (e.g., Harris, 1985; Clerkin et al., 1995; Kim et al., 2000; Gorunescu et al., 2002; Kim and Horowitz, 2002; McManus et al., 2004).

The ability to cope with uncertainty is considered to be an essential part of modern health care scheduling. Litvak and Long (2000) argue that, although an important source of the variability in hospitals is due to the urgent (non-elective) cases, a large part of the variability can be reduced by intelligent scheduling of the elective cases. Several authors have attempted to manage uncertainty by anticipating on the interaction of the planned (elective) and the emergency (non-elective) cases. Gerchak et al. (1996) provide a stochastic dynamic programming model for the advance scheduling of elective surgery under uncertain demand for emergency surgery. Bowers and Mould (2004) propose a policy of including planned, elective patients within the trauma (non-elective) session and show by means of simulation how substantially greater throughputs can be achieved. Kim et al. (2000) describe a flexible bed allocation scheme that reserves one or more beds for the exclusive use of elective-surgery patients to enhance the operations of the intensive care unit. Kim and Horowitz (2002) elaborate on this work and show through a simulation model that the combination of this flexible bed allocation scheme and a quota system for elective surgery greatly reduces the number of canceled surgeries.

We model uncertainty by means of probabilistic distribution functions and optimize expected performance. This way of dealing with scheduling under uncertainty is often referred to as stochastic scheduling. Alternative ways of coping with uncertainty include fuzzy scheduling and robust scheduling. An in-depth discussion on the differences between stochastic and fuzzy approaches to multi-objective mathematical programming under uncertainty can be found in Slowinski and Teghem (1990). In fuzzy scheduling, uncertainty is modeled using the concept of so-called fuzzy sets. A fuzzy set is characterized by a membership function, which maps the members of the universe into the unit interval $[0,1]$. The value 0 means that the member is not included in the given set, 1 describes a fully included member. The values between 0 and 1 characterize fuzzy members. In the context of uncertainty, such a membership function models the statement of how probable it is for a certain event to occur. A quality exposition on scheduling under fuzziness is provided by Slowinski and Hapke (2000).

Robust or proactive scheduling is concerned with building schedules that are protected against the occurrence of unexpected events. A robust schedule is able to absorb some level of unforeseen events without rescheduling. Accordingly, robust scheduling aims at maximizing the stability of the schedule. A review on robust scheduling techniques for project scheduling can be found in Herroelen and Leus (2004). Hans et al. (2005) propose several constructive and local search heuristics for the robust surgery loading problem. The objective is to assign the surgeries by the specialties in such a way, that the risk of working in overtime is minimized, no surgeries are canceled, and at the same time the operating room capacity utilization can be improved.

The remainder of this paper is structured as follows. Section 2 gives an outline of the theoretical background of the models applied in this study. Section 3 gives some more information on the hospital that has provided the data for this case study. Section 4 contains a discussion of the input analysis. More specifically, it is explained how the multinomial probability distributions are fitted for both the number of operated patients and the LOS of each operated patient for each surgeon-hospitalization unit combination. Section 5 contains a presentation of the graphical user interface that was built on top

of the algorithms to visualize the operation and performance of the system. Section 6 discusses the results obtained by applying the different approaches while Section 7 draws conclusions and lists some topics for future research.

2 Theoretical background

The algorithms that are applied in this paper to build the master surgery schedule can be divided into two classes. The first class consists of mixed integer programming (MIP) approaches. A distinction is made between linear min-max optimization approaches and a quadratic optimization approach. The second class consists of a simulated annealing heuristic that minimizes the total probability of bed shortage or, alternatively, the total expected bed shortage. The general principle behind both the MIP approaches and the metaheuristic approach is the same. Using the information on the stochastic distribution functions of the number of operated patients per block as well as the length of stay of each operated patient, both the mean and the variance of the daily bed occupancies by the elective cases can be calculated exactly for each hospitalization unit. To this purpose, the contribution of each surgeon-block allocation to the mean and variance of the daily bed occupancy of each hospitalization unit must be known. The respective formulas to calculate these contributions have been derived in Beliën and Demeulemeester (2005).

The way in which the mean and variance of the daily bed occupancies are used to build a good cyclic master surgery schedule differs between the different approaches. In the linear MIP approaches, the maximum of the daily mean (variance of the) bed occupancies is minimized. In the quadratic MIP approach, the daily mean (variance of the) occupancies are explicitly leveled by minimizing the quadratic sum. The MIP models are stated in Appendix A. The models are solved by a state-of-the-art mixed integer programming optimizer.

For the simulated annealing approach, shortage probabilities are calculated by assuming normally distributed bed occupancies, making use of the central limit theorem.

Additionally, by applying numerical integration techniques, expected shortages can be calculated. The objective function then involves the minimization of either the total shortage probability or the total expected shortage. To achieve this objective the algorithm iteratively explores neighbor solutions. A neighborhood move either involves an exchange of full block allocations (this move may include several surgeons) or involves an exchange between individual surgeon allocations to blocks. If an exchange leads to a better solution, the change is accepted. Otherwise, the change is rejected with an increasing probability towards the end of the search process. A more detailed description of the simulated annealing procedure is presented in Appendix B.

3 Case study

The case study presented entails the Virga Jesse Hospital, situated in Hasselt, Belgium. The 2004 annual report of this medium-sized hospital shows an important increase in activities. In 2004 the number of inpatient admissions has grown to a historical record of 21,923. In the same year the total revenues increased with 13 million Euro up to 163 million Euro. Also the number of outpatient admissions (referred to as day hospitalizations) has risen, while the average length of stay per patient has decreased. The election of the Belgian HR manager of the year and the laureate of the prestigious Belgian Tyco health care price for excellence in health care management are the evidences of Virga Jesse's top-class service.

Virga Jesse's central operating room complex consists of 9 rooms in which a total of 46 surgeons have been assigned operating room time. These surgeons are classified into 15 different surgical groups with respect to the specialism. Each operating room is open from Monday to Friday for 8.5 hours. Up to now, no elective surgery takes place during the weekends. The operated patients recover in one of the 25 hospitalization units of which only 10 units have served more than 100 elective cases in 2004. The models applied in this study involve the development of a (cyclic) master surgery schedule with leveled bed occupancy in these 10 major hospitalization units.

Table 1: Snapshot of the input file containing detailed information on all surgical interventions in 2004

| OR_NR | SURGEON | ROOM | HOSP. UNIT | DATE_IN | DATE_OUT |
|----------|---------|-----------------|------------|----------------|-----------------|
| 23005838 | PUTE | Operatiezaal 04 | 3200 | 2/01/2004 8:00 | 2/01/2004 17:00 |
| 23116828 | DTRG | Operatiezaal 09 | 3200 | 2/01/2004 8:00 | 2/01/2004 17:00 |
| 23408780 | VDVG | Operatiezaal 03 | 2150 | 2/01/2004 8:00 | 5/01/2004 15:00 |
| 23409553 | BOES | Operatiezaal 05 | 2160 | 2/01/2004 8:00 | 5/01/2004 15:19 |
| 23382108 | PUTE | Operatiezaal 04 | 3200 | 2/01/2004 8:05 | 2/01/2004 17:00 |
| 23383582 | LENH | Operatiezaal 08 | 3200 | 2/01/2004 8:05 | 2/01/2004 17:00 |
| 23409151 | PUTE | Operatiezaal 04 | 3200 | 2/01/2004 8:10 | 2/01/2004 17:00 |
| 23408550 | PUTE | Operatiezaal 04 | 3200 | 2/01/2004 8:15 | 2/01/2004 17:00 |
| 23382105 | PUTE | Operatiezaal 04 | 3200 | 2/01/2004 8:20 | 2/01/2004 17:00 |
| 23408576 | VDKJ | Operatiezaal 06 | 3200 | 2/01/2004 8:20 | 2/01/2004 17:00 |
| ... | | | | | |

4 Input analysis

Both the MIP based approaches and the simulated annealing approach require as input for each surgeon-hospitalization unit combination the probability distributions of the number of patients per block and the LOS for each operated patient. The theoretical models assume multinomial distributions, often referred to as empirical discrete probability distributions. These general probability distributions can easily be constructed from a database containing the detailed information on all surgical interventions that have been performed in a reasonably long time period (e.g., one year). Table 1 contains a snapshot of the (relevant) fields of the input file.

A procedure has been written that reads in these data records provided as an ASCII text file and automatically constructs the probability distributions for both the number of patients per block and the LOS per patient for each surgeon-hospitalization unit combination. This procedure simply counts the number of cases on each day for each surgeon-hospitalization unit combination. When these numbers are divided by the total number of respective surgery days, the probabilities for the number of operated patients per block that recover in the corresponding hospitalization unit are obtained. Since totals are made per day, it is implicitly assumed that a surgeon cannot be assigned to more than one block per day. This assumption holds in our case study as well as for many other hospitals. The same reasoning is applied for constructing the LOS distributions,

Table 2: Example of nr. patient and LOS distributions for three hospitalization units for surgeon DUPA

| SURGEON | HOSP. UNIT | NR. PATIENTS | | LOS | |
|---------|------------|--------------|-------|----------|-------|
| | | NR. PAT. | PROB. | NR. DAYS | PROB. |
| DUPA | 2160 | 0 | 0.20 | 3 | 0.20 |
| | | 1 | 0.38 | 4 | 0.02 |
| | | 2 | 0.34 | 5 | 0.02 |
| | | 3 | 0.06 | 6 | 0.03 |
| | | 4 | 0.02 | 7 | 0.28 |
| | | | | 8 | 0.21 |
| | | | | 9 | 0.21 |
| | | | | 10 | 0.03 |
| | | | | 12 | 0.02 |
| | 2601 | 0 | 0.56 | 4 | 0.03 |
| | | 1 | 0.34 | 7 | 0.04 |
| | | 2 | 0.06 | 8 | 0.41 |
| | | 3 | 0.04 | 9 | 0.45 |
| | | | | 10 | 0.07 |
| | 3200 | 0 | 0.16 | 1 | 1 |
| | | 1 | 0.10 | | |
| | | 2 | 0.22 | | |
| | | 3 | 0.30 | | |
| | | 4 | 0.12 | | |
| | | 5 | 0.08 | | |
| | | 6 | 0.02 | | |

but now obviously no intermediate day totals have to be made.

Only elective (planned) interventions are taken into account. The reason why the non-elective (emergency) cases are not retained is twofold. First, the occurrence as well as the recovery period of non-elective, emergency cases is, by definition, highly unpredictable and hence it would make little or no sense to fit a probability distribution to them. Second, non-elective cases often take place in blocks not preserved for the surgeon performing the surgery. Taking them into account would lead to a biased distribution for the number of patients per operating room block.

Table 2 shows an example of the output of this procedure, i.e., the derived probability distributions, for one particular surgeon. It must be clear at this point that the LOS distributions are specific for each surgeon-hospitalization unit combination. This is a very realistic basic assumption since the patient recovery time is usually strongly

related to this unique combination as patients operated by the same surgeon and recovering in the same hospitalization unit often suffer from similar ailments. Of course, surgeons can perform different surgical treatments in one block, but the proportions of these treatments are often reasonably constant.

Before applying this procedure, the surgeons and the existing schedule have to be read in manually. The existing schedule is needed to determine whether the case is elective or non-elective. If the intervention takes place on a day during which a block is preserved for the surgeon, it is considered to be an elective case. Otherwise, it is considered to be a non-elective case. A problem arises when a surgeon is assigned to more blocks having different durations. In this case, a ‘dummy’ surgeon is introduced for each different block duration. For instance, consider a surgeon who has been assigned one block of 8.5 hours on Monday and one block of 4 hours on Tuesday. In our approach, distributions will be derived for the Monday block as well as for the Tuesday block by introducing a ‘dummy’ surgeon for the latter. This implies that block durations are considered to be fixed when searching for better schedules. It also implies that hours cannot be exchanged between blocks. Only shifting of total blocks will be allowed.

The choice for this approach is justified as follows. First of all, a ‘block’ is probably the best unit for deriving the probability distributions. A smaller unit (e.g., an hour) is in our view less effective to fit the real distributions. Second, a block that extends twice as long as another block, assigned to the same surgeon, does not necessarily include twice the number of patients. Hence, not introducing a dummy surgeon would lead to derived probability distributions basically representing a mixture of two or more distributions. Third, working with fixed block durations entails some interesting computational features. It enables us to *a priori* calculate the per surgeon bed occupancy contributions. These contributions are needed as input for the mathematical programming models. With variable block sizes on the other hand, one could only calculate these contributions when the number of hours assigned per block is known. This would dramatically complicate the problem. Fourth, the graphical user interface is kept extremely simple as block assignments and exchanges can easily be done by dragging and

dropping. Finally, from a practical point of view, most of the surgeons have no different block durations and hence relatively few dummy surgeons have to be introduced.

5 Graphical user interface

In this section the graphical user interface (GUI) is presented. The application was programmed in Visual C++.NET and calls the CPLEX MIP solver (ILOG, 2002) for linear and quadratic optimization. The GUI visualizes the surgery schedule and the resulting bed usage occupancy distributions for a given schedule. Moreover, it allows the user to modify an existing schedule and to view the impact of a change in the schedule on the bed occupancy. Data like the schedule properties, the surgeon properties and the hospitalization properties can easily be read in and modified. Automation features include the deduction of the probability distributions for patient numbers and lengths of stay from a database (as described in Section 4) and the optimization of the schedule with respect to certain objective measures. Figure 1 shows an overview of the GUI with an empty surgery schedule.

The main window is divided into four views. In the upper left pane the (empty) master surgery schedule is shown. The seven columns in the grid represent the seven days of the week. The nine rows represent the nine operating rooms. Each room is open on each weekday for 8.5 hours. A subset of the surgeons is shown above the grid. The schedule could now be built easily from scratch by dragging and dropping the surgeons to the timetable cells. Of course, a room can also be assigned for a limited number of hours instead of the full 8.5 hours. Each assignment introduces a patient flow in the system, which is reflected by an increase in the bed occupancy of one or more hospitalization units on one or more days. This is represented in the upper right pane. Schedules could also be built automatically while aiming at certain optimization objectives. The computational results (like solution time, solution quality, etc.) are given in the lower left pane. Finally, the right bottom pane is a simulation pane. A simulation run could be done in order to validate the theoretical basic assumptions (mainly the central limit theorem) of the model. To this purpose it can be verified whether the predicted bed

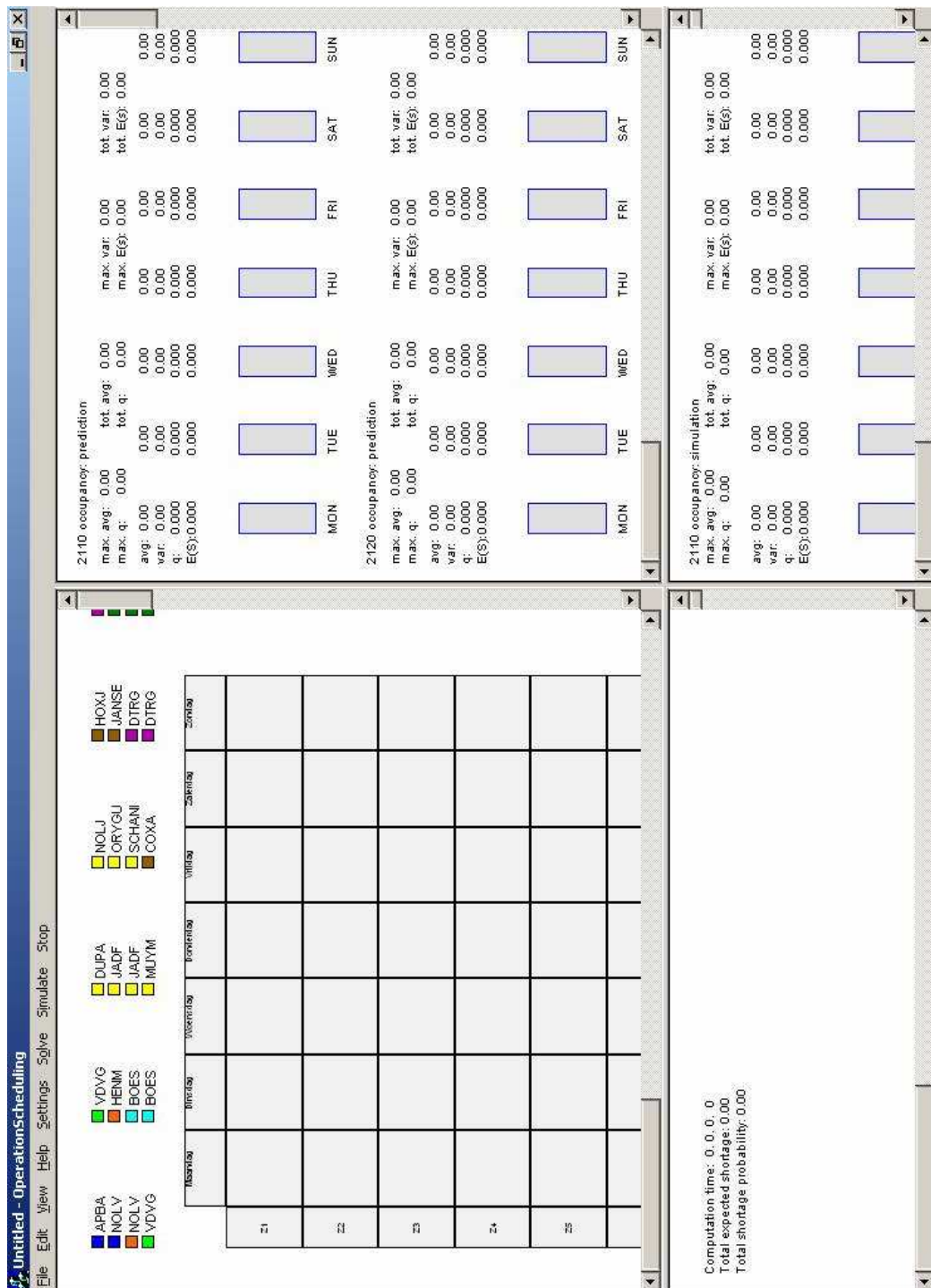


Figure 1: Overview of the GUI with empty schedule

occupancies (and shortages) obtained by calculation are similar to the ones obtained by simulation. Figure 2 shows the current master surgery schedule with resulting bed occupancy (only three hospitalization units are shown). The small T-ending bars on top of each colored occupancy box indicate the standard deviations of the bed occupancy distributions on the corresponding days at the corresponding hospitalization units.

Using dialog boxes, the schedule, surgeon and hospitalization unit properties could easily be modified. As an example some of the dialog boxes for editing the surgeon properties are represented in Figure 3. The left dialog box shows the surgeon basic properties and a list of the hospitalization units to which patients of the selected surgeons flow. The user can select one of these units to edit. The upper right dialog box then allows the user to choose between the number of patients distribution or the LOS distribution for editing. The lower right dialog box finally allows the user to edit individual distribution values (number and probability) of the LOS distribution for this particular surgeon-hospitalization unit combination.

Concerning the automation procedures, one basically can choose between two approaches: a mixed integer programming procedure, either a linear or a quadratic one, that aims at leveling the bed occupancy of one or more hospitalization units, or a simulated annealing approach that directly tries to minimize the total shortage probability or the total expected shortage. To provide additional details for the integer programming procedure, the dialog box shown in Figure 4 is displayed. The user can choose between the linear or the quadratic variant, specify the maximum running time limit and provide the objective function weights of the respective hospitalization units (2110, 2120, 2130, ...) for the mean (first column) as well as the variance (second column). The weights represent the relative importance of the leveled bed occupancy for the respective hospitalization units; e.g., a weight of 1 (0) indicates that a particular hospitalization unit is (not) taken into account.

When the maximal time limit is reached, the dialog box displayed in Figure 5 pops up. The user obtains information about the optimality status and can choose either to

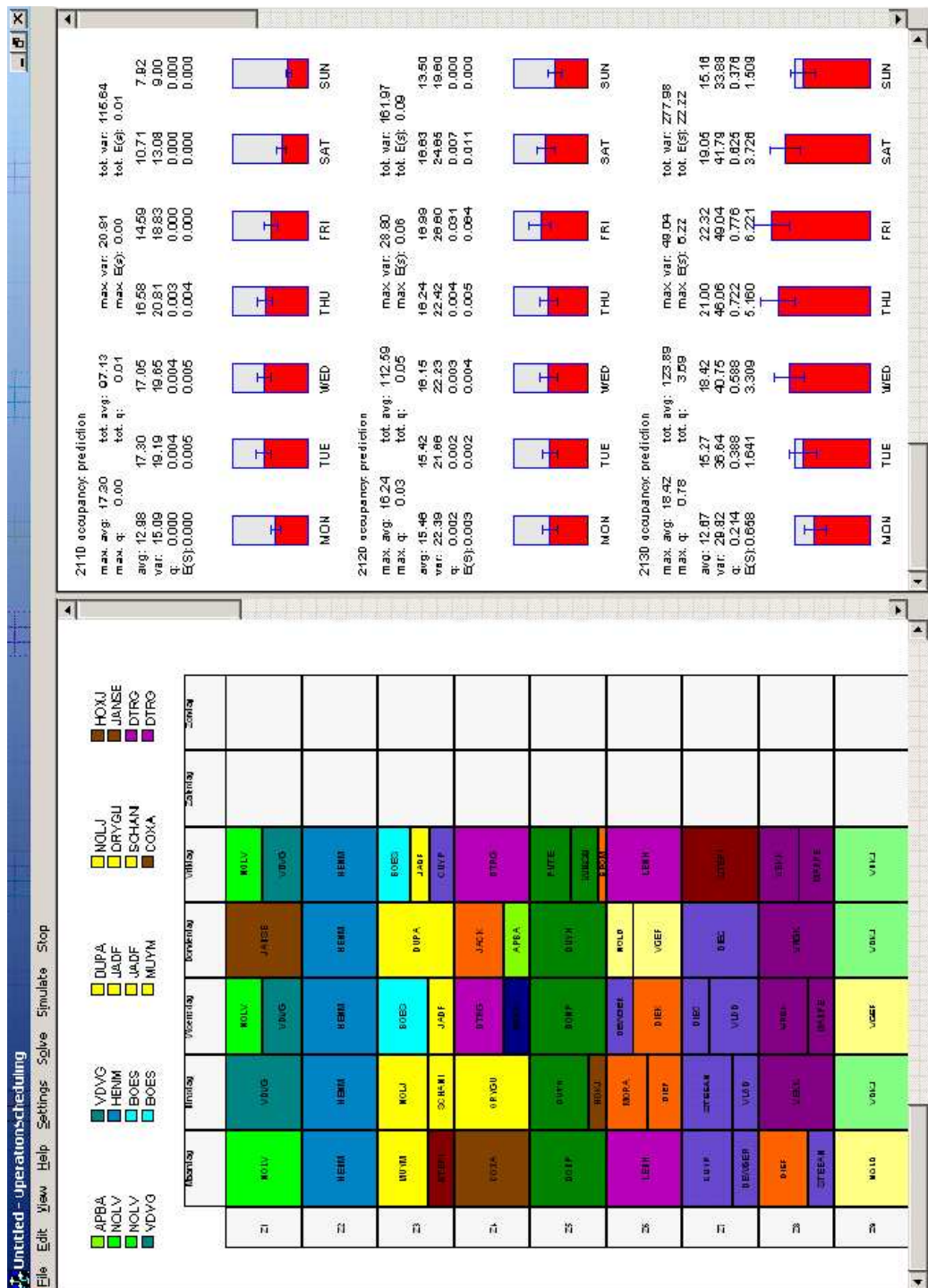


Figure 2: Current master surgery schedule with resulting bed occupancy (only three hospitalization units shown)

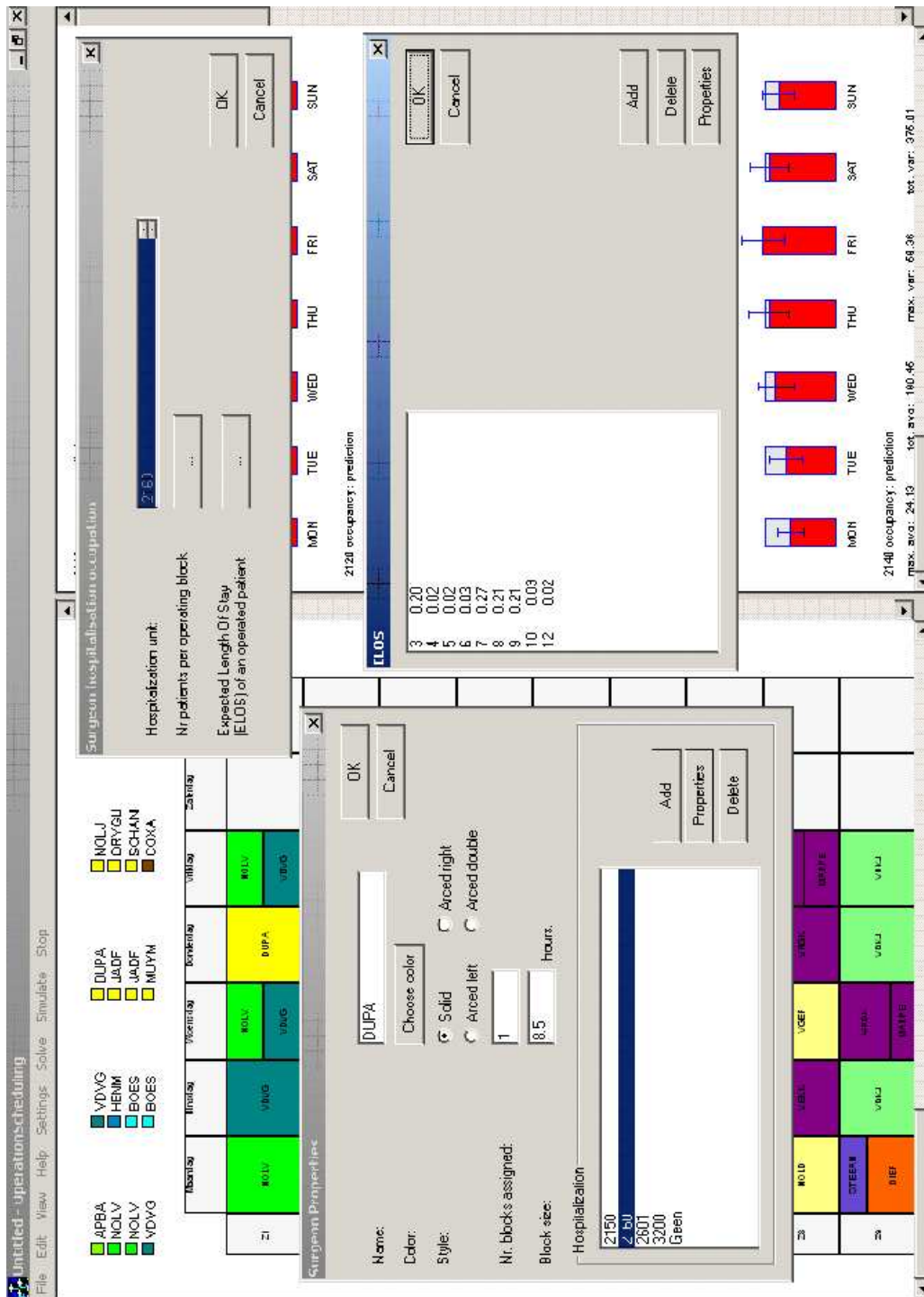


Figure 3: Editing the properties of a surgeon

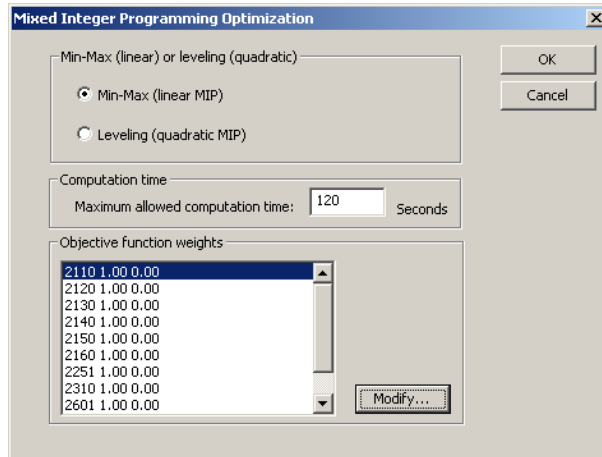


Figure 4: Dialog box: Starting a MIP

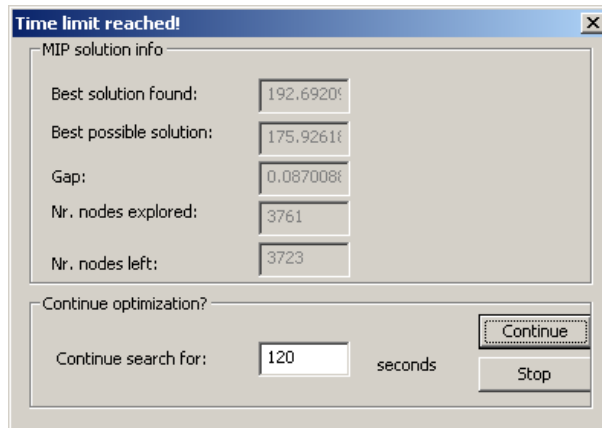


Figure 5: Dialog box: Upon completion of the MIP

stop the algorithm or to continue the search for an additional time span. The given information includes the current best found objective value, the lower bound, the gap, that is the difference between the current solution and the lower bound expressed as a percentage of the second, and the number of explored and non-explored nodes in the search tree.

To start a simulated annealing (SA) procedure, the user has to specify a number of general SA settings (initial temperature, temperature update interval and update factor), give the stop criteria (max. nr. iterations, max. time limit), indicate the

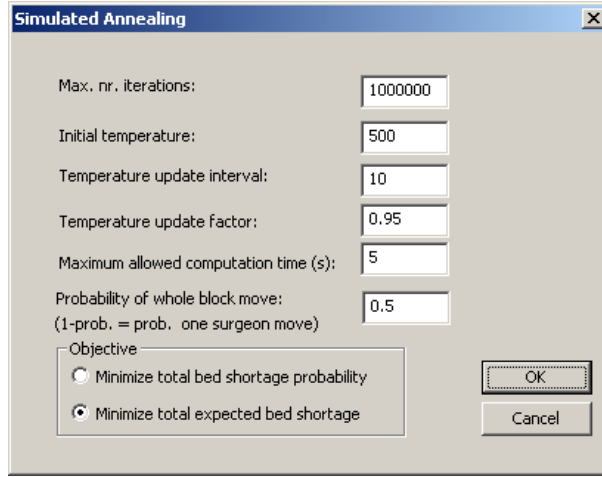


Figure 6: Dialog box: Starting a simulated annealing procedure

probability of a whole block move (which automatically determines the probability of a one surgeon move) and choose between the two objective functions. The dialog box to provide this information is shown in Figure 6.

6 Results

Rather than trying to find the overall best master surgery schedule for the Virga Jesse Hospital, which is a subjective matter after all and hence makes little sense anyway, we discuss and compare the results of a number of different algorithm runs.

As can be seen in Figure 2, some problems may arise at hospitalization unit 2130 (third unit, shown at the bottom), where there is a high peak occupancy on Friday leading to a positive expected bed shortage. An optimization procedure that exclusively focuses on this hospitalization unit could turn out to be useful to solve the problem. The resulting schedule of a linear MIP, aiming at the minimization of the maximum mean occupancy peak of unit 2130 is shown in Figure 7. It should be clear that the bed occupancy in unit 2130 is now much more leveled over the week.

Also hospitalization unit 2140 suffers from large differences in the bed occupancy peaks. This asks for a scheduling procedure that simultaneously focusses on the leveling

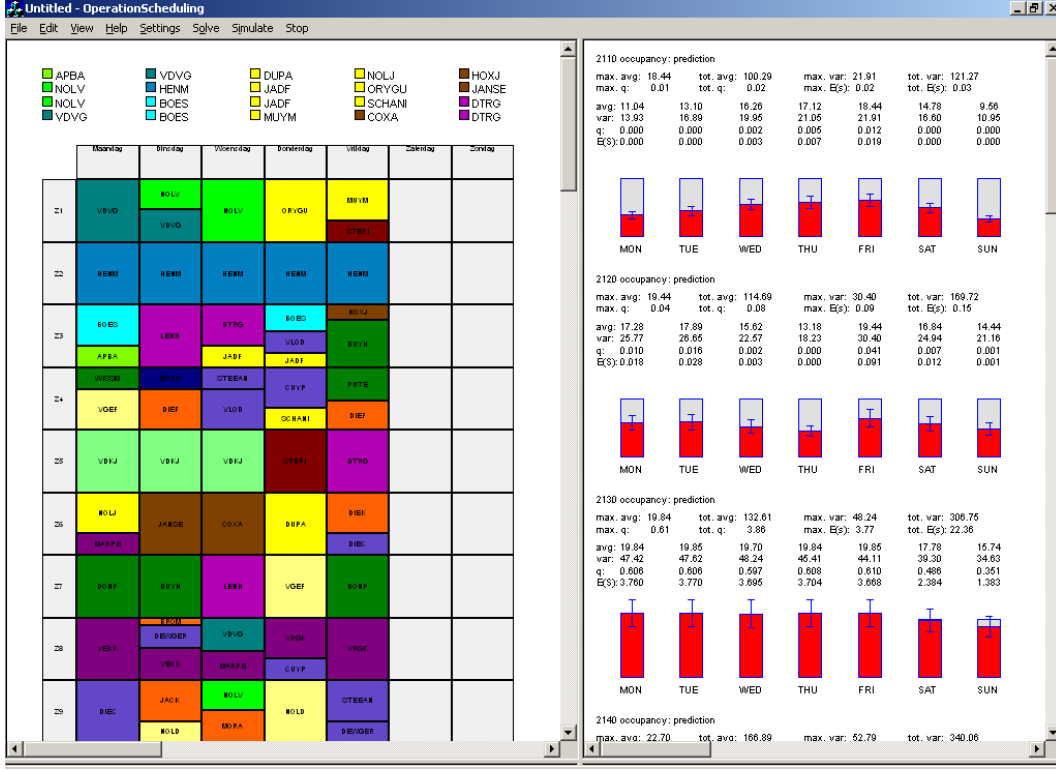


Figure 7: The results of a linear MIP to level the mean bed occupancy of hospitalization unit 2130 (shown in the lower right)

of the bed occupancy distributions in units 2130 and 2140. To this purpose, a linear MIP procedure that minimizes the weighted maximum peak of the bed occupancies in units 2130 and 2140 could be applied. However, as we already presented a solution based on a linear MIP procedure (see Figure 7), we present the results of a quadratic MIP with weights 1 for units 2130 and 2140 and 0 for all other units. Figure 8 contains the resulting schedule.

Finally, Figure 9 displays the schedule that results from applying a simulated annealing procedure with the settings shown in Figure 6. This procedure tries to minimize the total expected bed shortage, taking all hospitalization units into consideration.

It is difficult to objectively compare the quality of the generated schedules, as there is no once and for all objective measure to make this comparison. To build a quality

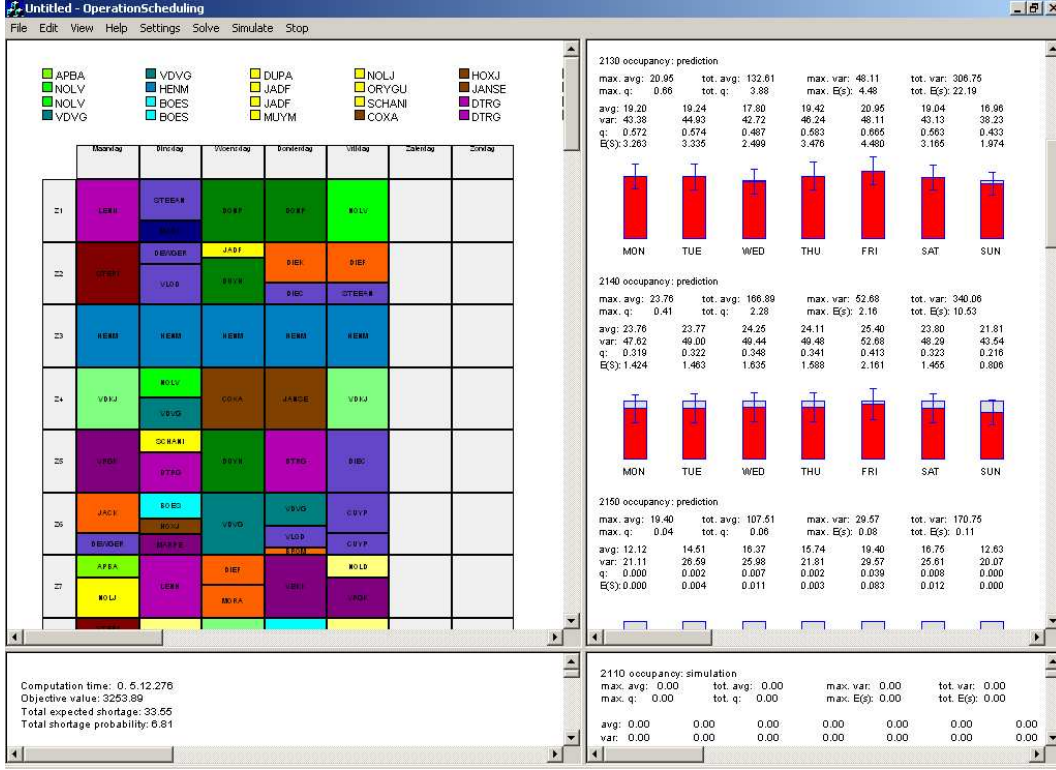


Figure 8: The results of a quadratic MIP to level the mean bed occupancy of hospitalization units 2130 and 2140

schedule or at least to improve the current schedule, one has to study the current practices and determine the most appropriate objective function and automation procedure. For instance, if capacity problems always occur at the same hospitalization unit, a linear or quadratic MIP procedure that focusses on this unit will probably render the best results. The visualization of the bed occupancies can of course assist in determining the appropriate model. However, there might be a different explanation for the variability in these occupancies rather than the variability in surgery admissions and LOS.

Suppose we take the expected number of bed shortages over all ten hospitalization units included in this study as the one and only objective measure. The results of several optimization procedures are shown in Table 3. The first line indicates the total expected bed shortage in the current schedule. This number (37.82) indicates that, over all hospitalization units, more than 5 beds per day are lacking in the assigned hospi-

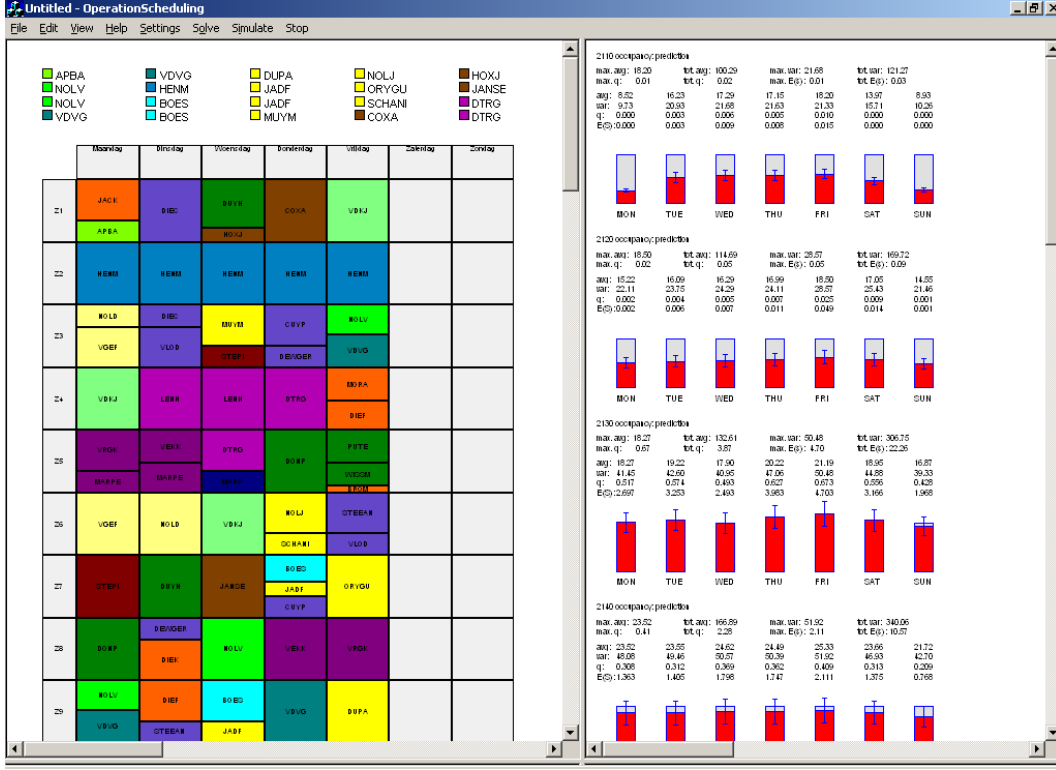


Figure 9: The results of a simulated annealing algorithm that minimizes the total expected bed shortage

talization unit and hence have to be found in another hospitalization unit. A possible explanation for this remarkably high number is as follows. Recall that we derived the per surgeon probability distributions from the daily records of a database containing all surgical interventions during a 1-year period. The resulting variability in these derived distributions is, however, probably higher than the real-life variability. Indeed, in real life it is probably the case that a surgeon admits or rejects patients as a function of the remaining bed capacity at the relevant hospitalization unit at that moment. In other words, an important part of the variability can be taken care of by appropriate admission of elective cases during the third stage of the surgery scheduling process, which involves the detailed planning of the individual elective cases in the allocated blocks. Obviously, in the concern of both patient and surgeon the postponement of surgery is best avoided as much as possible. Therefore, methods for a careful design of the master surgery schedule, as presented in this study, are still valuable.

Table 3: Minimizing the total expected bed shortage

| Procedure | Total exp. shortage | Comp. time (s) |
|-----------------------------------|---------------------|----------------|
| Current schedule | 37.82 | - |
| Lin. MIP MIN-MAX mean 2130 | 34.44 | 120 |
| Lin. MIP MIN-MAX mean + var. 2130 | 34.12 | 120 |
| Quad. MIP mean 2130, 2140 | 33.55 | 120 |
| Quad. MIP mean all units | 33.29 | 360 |
| SA | 33.12 | 120 |

Table 3 shows that the total expected bed shortage drops from 37.82 to 34.44 if a linear MIP approach is used in which the maximal mean bed occupancy of the bottleneck hospitalization unit 2130 is used. The shortage decreases further to 34.12 if also the variance of the bed occupancy in this unit is taken into consideration. If a quadratic MIP procedure is applied on the two most heavily loaded units 2130 and 2140, the expected shortage decreases to 33.55 and if all units are taken into account to 33.29. Finally, application of a simulated annealing procedure that directly aims at this objective results in a schedule with expected bed shortage equal to 33.12.

It must be clear that this result does not necessarily mean that the schedule displayed in Figure 9 is the best schedule for Virga Jesse. It is just a possible schedule that has the best score (at least amongst the schedules resulting from the few procedures we have tested) on one measure. The real power of the software lies in the visualization of the schedule and the resulting bed occupancy, the ease with which schedules can be built and the capability it provides to carry out an in-depth analysis of the existing system. Using the software, managers can find answers to questions like “what is the most leveled bed occupancy possible at hospitalization unit X?” or “which schedule simultaneously levels the bed occupancy in units X and Y?”.

7 Conclusions and future research

This paper has presented a case study involving an application of the models developed in Beliën and Demeulemeester (2005) to build a master surgery schedule with leveled resulting bed occupancy. To this purpose, the required input data, namely the distribu-

tion functions for the number of operated patients as well as for the length of stays, have been derived from the central database containing detailed information on all surgical cases during a 1-year period in a medium-sized Belgian hospital. The graphical user interface hides the algorithmic procedures and makes it easy to build a schedule having particular features like the most leveled bed occupancy in a certain hospitalization unit or the smallest overall bed shortage probability. Depending on the hospital's situation, and in particular on the problems it is facing, a procedure can be chosen to build a new master surgery schedule. Additionally, the application can provide managers with important insights into the behavior of the system.

The models make abstraction of the differences between the operating rooms, i.e., for the resulting bed occupancy it is completely irrelevant whether a surgeon is allocated to room 1 or to room 2. As a consequence, the models often result in a schedule in which surgeons from different disciplines have to share one operating room, sometimes even on the same day. However, in many cases surgeons of the same group prefer to be scheduled in the same room for this includes several practical benefits. Of course, shifting surgeons to other rooms within the same day has no impact upon the bed occupancies. Hence, the scheduler can easily swap room allocations within the same day in order to group surgeons having a similar discipline into the same room as much as possible. Instead of doing this manually, a post-improvement heuristic could be written to do the job. Alternatively, room restrictions could be an integral part of the optimization model, either in the objective function or as additional constraints. This is an interesting direction for further research.

A second shortcoming of the proposed models is the assumption that the generated schedule is completely repeated each cycle time. It might be more efficient for those surgeons having assigned a small number of hours of operating time per cycle to aggregate these hours and only operate once in two or three cycles. Obviously, a prolongation of the cycle time would deal with this issue. However, surgeons also like to have their schedule as simple as possible which entails as few changes as possible from week to week. The latter is not guaranteed if we simply apply the models described above with

a longer cycle time. Blake et al. (2002) overcome this problem as follows. In a first phase, they relax the surgery demand constraints in their integer programming model, taking as an objective for their cycle master surgery schedule the minimization of the undersupply of target operating room hours (see also Blake and Donald, 2002). In a second phase, a post-improvement heuristic is run that tries to further improve this objective by introducing some changes in the schedule from week to week. A similar heuristic could be written to accompany our cyclic surgery scheduling models. Alternatively, we could work the other way around; i.e., constructing a schedule with a longer cycle time (e.g., two or three weeks) and afterwards running an improvement algorithm (either heuristic or exact) that tries to minimize the changes in this schedule from week to week with none (or few) repercussions for our objective function.

Acknowledgements

We acknowledge the support given to this project by the Fonds voor Wetenschappelijk Onderzoek (FWO) - Vlaanderen, Belgium under contract number G.0463.04. We are grateful to Jurgen Huygh (Virga Jesse Hospital, Hasselt, Belgium) for his support to this project and to Geert Moechars (Virga Jesse Hospital, Hasselt, Belgium) for providing the case study data.

References

- Beliën, J. and Demeulemeester, E. (2005). Building cyclic master surgery schedules with leveled resulting bed occupancy. To appear in *European Journal of Operational Research*.
- Blake, J. T. and Carter, M. W. (2002). A goal programming approach to strategic resource allocation in acute care hospitals, *European Journal of Operational Research* **140**: 541–561.
- Blake, J. T. and Carter, M. W. (2003). Physician and hospital funding options in a public system with decreasing resources, *Socio-Economic Planning Sciences* **37**: 45–68.

- Blake, J. T., Dexter, F. and Donald, J. (2002). Operating room manager’s use of integer programming for assigning block time to surgical groups: A case study, *Anesthesia and Analgesia* **94**: 143–148.
- Blake, J. T. and Donald, J. (2002). Mount Sinai hospital uses integer programming to allocate operating room time, *Interfaces* **32**: 63–73.
- Bowers, J. and Mould, G. (2004). Managing uncertainty in orthopaedic trauma theatres, *European Journal of Operational Research* **154**: 599–608.
- Clerkin, D., Fos, P. J. and Petry, F. E. (1995). A decision-support system for hospital bed assignment, *Hospital and Health Services Administration* **40**: 386–400.
- Dexter, F., Macario, A. and O’Neill, L. (2000). Scheduling surgical cases into overflow block time - Computer simulation of the effects of scheduling strategies on operating room labor costs, *Anesthesia and Analgesia* **90**: 980–988.
- Dexter, F., Macario, A. and Traub, R. D. (1999). Which algorithm for scheduling add-on elective cases maximizes operating room utilization?, *Anesthesiology* **91**: 1491–1500.
- Dexter, F. and Traub, R. D. (2002). How to schedule elective surgical cases into specific operating rooms to maximize the efficiency of use of operating room time, *Anesthesia and Analgesia* **94**: 933–942.
- Dexter, F., Traub, R. D. and Lebowitz, P. (2001). Scheduling a delay between different surgeons’ cases in the same operating room on the same day using upper prediction bounds for case durations, *Anesthesia and Analgesia* **92**: 943–946.
- Galassi, M., Davies, J., Theiler, J., Gough, B., Jungman, G., Booth, M. and Rossi, F. (2003). *GNU Scientific Library Reference Manual Edition 1.3*, Network Theory LTd.
- Gerchak, Y., Guptar, D. and Henig, M. (1996). Reservation planning for elective surgery under uncertain demand for emergency surgery, *Management Science* **42**: 321–334.

- Gorunescu, F., McClean, S. I. and Millard, P. H. (2002). A queueing model for bed-occupancy management and planning of hospitals, *Journal of the Operational Research Society* **53**: 19–24.
- Guinet, A. and Chaabane, S. (2003). Operating theatre planning, *International Journal of Production Economics* **85**: 69–81.
- Hans, E. W., Wullink, G., van Houdenhoven, M. and Kazemier, G. (2005). Robust surgery loading, *Technical Report Beta-wp141*, dep. Operational Methods for Production and Logistics, University of Twente.
- Harris, R. A. (1985). Hospital bed requirements planning, *European Journal of Operational Research* **25**: 121–136.
- Herroelen, W. and Leus, R. (2004). Robust and reactive project scheduling: A review and classification of procedures, *International Journal of Production Research* **42**(8): 1599–1620.
- Hughes, W. L. and Soliman, S. Y. (1985). Short-term case mix management with linear programming, *Hospital and Health Services Administration* **30**: 52–60.
- ILOG (2002). *ILOG CPLEX 8.1 User's Manual*.
- Jebali, A., Alouane, A. B. H. and Ladet, P. (2006). Operating rooms scheduling, *International Journal of Production Economics* **99**: 52–62.
- Kim, S.-C. and Horowitz, I. (2002). Scheduling hospital services: The efficacy of elective-surgery quotas, *Omega - the International Journal of Management Science* **30**: 335–346.
- Kim, S.-C., Horowitz, I. and Buckley, T. A. (2000). Flexible bed allocation and performance in the intensive care unit, *Journal of Operations Management* **18**: 427–443.
- Lapierre, S. D., Batson, C. and McCaskey, S. (1999). Improving on-time performance in health care organizations: A case study, *Health Care Management Science* **2**: 27–34.
- Litvak, E. and Long, M. C. (2000). Cost and quality under managed care: Irreconcilable differences?, *The American Journal of Managed Care* **6**: 305–312.

- Marcon, E., Kharraja, S. and Simonnet, G. (2003). The operating theatre planning by the follow-up of the risk of no realization, *International Journal of Production Economics* **85**: 83–90.
- McManus, M. L., Long, M. C., Cooper, A. and Litvak, E. (2004). Queuing theory accurately models the need for critical care resources, *Anesthesiology* **100**: 1271–1276.
- Ozkarahan, I. (1995). Allocation of surgical procedures to operating rooms, *Journal of Medical Systems* **19**(4): 333–352.
- Ozkarahan, I. (2000). Allocation of surgeries to operating rooms using goal programming, *Journal of Medical Systems* **24**(6): 339–378.
- Rifai, A. K. and Pecenka, J. O. (1989). An application of goal programming in healthcare planning, *International Journal of Production Management* **10**: 28–37.
- Robbins, W. A. and Tuntiwongbiboon, N. (1989). Linear programming is a useful tool in case-mix management, *Healthcare Financial Management* **43**: 114–116.
- Slowinski, R. and Hapke, M. (eds) (2000). *Scheduling Under Fuzziness*, Vol. 37 of *Studies in Fuzziness and Soft Computing*, Springer-Verlag.
- Slowinski, R. and Teghem, J. (eds) (1990). *Stochastic versus Fuzzy Approaches to Multiobjective Mathematical Programming under Uncertainty*, Kluwer Academic Publishers, Dordrecht.
- Weiss, E. N. (1990). Models for determining estimated start times and case orderings in hospital operating rooms, *IIE Transactions* **22**: 143–150.

APPENDIX A

In this Appendix we state the mixed integer programming models that are used to develop the master surgery schedules with leveled resulting bed occupancy. The notation used in these models is as follows:

The indices and sets are:

i, j, d, d_1, d_2 : days in the cycle

s : surgeons

r : rooms

h : hospitalization units

A : set of days in the cycle

S : set of surgeons

R : set of rooms

H : set of hospitalization units

The decision variable is:

$$x_{isr} = \begin{cases} 1, & \text{if surgeon } s \text{ obtains an operating room block in room } r \text{ on day } i \\ 0, & \text{otherwise} \end{cases}$$

The help variables are:

$mean_{hi}$ = the mean bed occupancy in hospitalization unit h on day i

\overline{mean}_h = the peak mean bed occupancy in hospitalization unit h over all days in the cycle

var_{hi} = the variance of the bed occupancy in hospitalization unit h on day i

\overline{var}_h = the peak variance of the bed occupancy in hospitalization unit h over all days in the cycle

The data parameters are:

| | |
|----------------|---|
| req_s : | the number of blocks required by surgeon s |
| cap_{ir} : | the total capacity (in hours) of room r on day i |
| dur_s : | the duration (in hours) of an assignment to surgeon s |
| l : | the length of the cycle time |
| $dist(i, j)$: | the ‘distance’ between day i and day j in the cycle $= i - j + 1$, if day j precedes day i and $= l + i - j + 1$, otherwise |
| p_{sdh} : | the probability that a patient operated by surgeon s recovers d days in hospitalization unit h |
| m_{sh} : | the maximal number of days a patient can stay in hospitalization unit h after surgery by surgeon s |
| n_{sh} : | the number of patients (per block) that recover in hospitalization unit h after surgery by surgeon s |

In order to avoid needlessly complicating the model, we only state the formulation for the case with deterministic patient numbers, i.e., a fixed n_{sh} for each surgeon-hospitalization unit combination. For the extension of the model to stochastic patient numbers, we refer to Beliën and Demeulemeester (2005). We start with the linear min-max model that aims at minimizing the weighted peaks in the expected bed occupancy and/or the variance in bed occupancy:

$$\text{Minimize } \sum_{h \in H} (w_{mean_h} \overline{mean}_h + w_{var_h} \overline{var}_h) \quad (1)$$

subject to:

$$\sum_{i \in A} \sum_{r \in R} x_{isr} = req_s \quad \forall s \in S \quad (2)$$

$$\sum_{s \in S} dur_s x_{isr} \leq cap_{ir} \quad \forall i \in A \text{ and } \forall r \in R \quad (3)$$

$$\sum_{r \in R} x_{isr} \leq 1 \quad \forall i \in A \text{ and } \forall s \in S \quad (4)$$

$$mean_{hi} = \sum_{s \in S} \sum_{j \in A} \sum_{r \in R} \left(\sum_{d=dist(i,j)}^{m_{sh}} p_{sdh} n_{sh} \lceil d/l \rceil \right) x_{jsr} \quad \forall h \in H \text{ and } \forall i \in A \quad (5)$$

$$\begin{aligned} var_{hi} = & \sum_{s \in S} \sum_{j \in A} \sum_{r \in R} \left(\sum_{d=dist(i,j)}^{m_{sh}} p_{sdh} (1 - p_{sdh}) n_{sh} \lceil d/l \rceil \right. \\ & \left. - \sum_{d_1=dist(i,j)}^{m_{sh}} \sum_{d_2=dist(i,j)}^{d_1-1} 2p_{sd_1h} p_{sd_2h} n_{sh} \lceil d_2/l \rceil \right) x_{jsr} \end{aligned} \quad \forall h \in H \text{ and } \forall i \in A \quad (6)$$

$$mean_{hi} \leq \overline{mean}_h \quad \forall h \in H \text{ and } \forall i \in A \quad (7)$$

$$var_{hi} \leq \overline{var}_h \quad \forall h \in H \text{ and } \forall i \in A \quad (8)$$

$$mean_{hi}, var_{hi} \geq 0 \quad \forall h \in H \text{ and } \forall i \in A \quad (9)$$

$$\overline{mean}_h, \overline{var}_h \geq 0 \quad \forall h \in H \quad (10)$$

$$x_{isr} \in \{0, 1\} \quad \forall i \in A, \forall s \in S \text{ and } \forall r \in R \quad (11)$$

The objective function (1) minimizes the weighted sum of peaks in the bed occupancy and variance over all hospitalization units. Constraint set (2) ensures that every surgeon obtains the number of blocks that were assigned by the case mix planning. Constraint set (3) makes sure that the total operating time assigned on each day in each room does not exceed the available operating room time. Constraint set (4) states that each surgeon gets at most one operating room block per day, i.e., it prevents a surgeon to be scheduled simultaneously in two different rooms. Constraint set (5) calculates the expected bed occupancy in each hospitalization unit as a function of the operating room schedule, while constraint set (6) calculates the variance on this occupancy. For the derivation of these expressions we again refer to Beliën and Demeulemeester (2005).

Constraint set (7) provides the link with the objective function by imposing for each hospitalization unit that the expected bed occupancy on each day cannot exceed the peak expected bed occupancy. Constraint set (8) does the same for the variance. Finally, constraint sets (9) and (10) define the mean and variance to be nonnegative and constraint set (11) defines x_{isr} as a binary decision variable.

The quadratic MIP model is identical to the linear min-max MIP model except for the objective function that represents an explicit (weighted) leveling of the mean and variance of the occupancy in the different hospitalization units:

$$\text{Minimize } \sum_{h \in H} (w_{mean_h} \sum_{i \in A} mean_{hi}^2 + w_{var_h} \sum_{i \in A} var_{hi}^2) \quad (12)$$

Obviously, constraints (7), (8) and (10) can now be removed as these are no longer required.

APPENDIX B

In this Appendix we briefly describe the simulated annealing algorithm that is used to build a cyclic master surgery schedule. We consider two possible objective functions. The first one minimizes the probability of a bed shortage, the second one minimizes the total expected bed shortage. To calculate these objective functions we employed a simplification of the model using the central limit theorem. According to this theorem, each variable which is the sum of a number of independent variables, is approximately normally distributed with mean equal to the sum of the independent means and variance equal to the sum of the independent variances. Recall that the independent means and variances can easily be calculated exactly using expressions (5) and (6). Hence, for calculating the shortage probabilities we can simply make use of the standard cumulative normal distribution functions. For calculating the expected shortages we have to apply numerical integration. For instance, to calculate the expected bed shortage (EB_{hi}) for hospitalization unit h on day i , we compute the following integral:

$$EB_{hi} \approx \int_{c_{hi}+0.5}^{+\infty} (z_{hi} - c_{hi}) \frac{1}{\sqrt{2\pi var_{hi}}} e^{-\frac{(z_{hi}-mean_{hi})^2}{2var_{hi}}} dz_{hi} \quad (13)$$

with z_{hi} a stochastic variable that represents the total number of occupied beds and c_{hi} the capacity of beds in hospitalization unit h on day i . This expression sums up all shortages ($z_{hi} - c_{hi}$) multiplied by the corresponding probabilities. The integral starts at $c_{hi} + 0.5$ (and not at c_{hi} or at $c_{hi} + 1$) to take into account a continuity correction for approaching a discrete function with a continuous one. These integrals were calculated by the numerical integration routines provided in *GNU Scientific Library* (GSL) version 1.3 (Galassi et al., 2003).

A basic SA implementation is used. We defined two different neighborhood moves. The first move swaps all surgeon assignments to a room on a particular day with all assignments within one room on a different day, provided that the available operating room time is not exceeded in one of the rooms. The second move only swaps one surgeon assignment, again taking the total time constraint into account. Each iteration in the SA procedure consists of one neighborhood move, the probability with which the

respective moves are chosen can be specified by the user. In each move, the first block is chosen randomly. The second block is the first encountered block for which a swap results in an improvement (decrease) of the objective value. If no such block can be found, the block leading to the smallest increase is chosen. Since swaps between the same surgeon and swaps within the same day have no impact on the objective function, these swaps are not taken into account. In order to decide whether or not to accept a worse solution, a standard Boltzman function is evaluated. Let T denote the temperature and Δf the decrease in objective function. For swaps with negative Δf the probability of acceptance is given by $e^{\frac{\Delta f}{T}}$. Of course, the best found schedule is saved.